(Knowledge for Development)

KIBABII UNIVERSITY

(KIBU)

UNIVERSITY EXAMINATIONS
2020/2021 ACADEMIC YEAR

END OF SEMESTER EXAMINATIONS
YEAR THREE SEMESTER TWO EXAMINATIONS

FOR THE DEGREE OF
BACHELORS OF SCIENCE
(INFORMATION TECHNOLOGY)

COURSE CODE : BIT 326

COURSE TITLE : DATA WAREHOUSING AND MINING

DATE: 12/10/2021 **TIME: 9.00 A.M-11.00 A.M**

INSTRUCTIONS TO CANDIDATES

ANSWER QUESTIONS ONE AND ANY OTHER TWO

QUESTION ONE [COMPULSORY] [30 MARKS]

a. Describe any two characteristics of a data warehouse. **[4 Marks]**

b. Identify any two different methods for information delivery. **[2 Marks]**

c. "A dimension table is wide and a fact table is deep". Elaborate. **[2 Marks]**

d. Using valid examples, explain the differences between type 2 slowly changing dimensions and type 3 slowly changing dimensions. **[3 Marks]**

e. Discuss any one reason why data quality critical in a data warehouse. **[2 Marks]**

f. Describe a factless fact table and identify a situation where a factless fact table may be used. **[3 Marks]**

g. Equity bank want to develop a data warehouse for effective decision-making about their loan schemes. The bank provides loans to customers for various purposes such as Car loan, educational Loan, Personal Loan, Business Loan etc. The bank has categorised the country into a number of regions namely; Eastern, Western, Central and Coast. Each region consists of a set of counties. Loan is disbursed to customers at interest rates that change from time to time. In addition, at any given point in time, the different types of loans have different rates. The data warehouse should record an entry for each disbursement of loan to a customer.

i. Design an information package diagram for the application. **[5 Marks]**

ii. Design a star schema for the data warehouse and clearly identify the fact table(s), dimension tables, their attributes and measures. **[5 Marks]**

iii. Describe an algorithm the bank can use to cluster its potential customers based on their attributes. **[2 Marks]**

iv. Describe how data warehousing and mining can help the bank increase its productivity. **[2 Marks]**

QUESTION TWO [20 MARKS]

a. Kibabii University wants to record marks for courses completed by students using dimensions: course, student, time and a measure average marks. Create a Cube with the following OLAP operations: **[10 Marks]**

i. Rollup

ii. Drill down

iii. Slice

iv. Dice

v. Pivot

b. With the aid of a diagram, discuss the typical architecture of a typical data mining system. **[10 Marks]**

QUESTION THREE [20 MARKS]

a. A manufacturing company has a huge sales network. To control the sales, it is divided into regions. Each region has multiple zones and each zone has different cities. Each sales person is allocated different cities. The objective is to track sales figure at different granularity levels of region and to count number of products sold.

i. Design a star schema by considering granularity levels for region, sales person and time. **[8 Marks]**

ii. Convert the star schema to snowflake schema. **[6 Marks]**

b. Discuss application and major issues in data mining. **[6 Marks]**

QUESTION FOUR [20 MARKS]

a. Explain why Metadata is important in data warehousing. **[2 Mark]**

b. Using examples to support your answers, discuss the types of metadata stored in a data warehouse. **[8 Marks]**

c. Discuss various OLAP models and their architecture. **[10 Marks]**

QUESTION FIVE [20 MARKS]

a. Describe the steps of the ETL (Extract-Transform-Load) cycle. **[6 Marks]**

b. Explain the role of Metadata in a data warehouse. **[6 Marks]**

c. A simple example from a Golf club involving only discrete ranges has Play Golf as categorical attribute with values (No, Yes) and a training data set shown below. Apply the decision tree algorithm and show the generated rules. **[8 Marks]**

Predictors				Target
Outlook	Temp.	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No